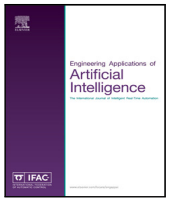
Engineering Applications of Artificial Intelligence 120 (2023) 105947

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Single-image HDR reconstruction by dual learning the camera imaging process

Lei She [a], Mao Ye [a,*], Shuai Li [b,c], Yu Zhao [a], Ce Zhu [c], Hu Wang [a]

[a] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, PR China
[b] School of Control Science and Engineering, Shandong University, Jinan 250000, PR China
[c] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, PR China

ARTICLE INFO

Keywords:
Dual learning
High dynamic range
Camera imaging process

ABSTRACT

It is a very challenging problem to reconstruct a high dynamic range (HDR) image from a single exposure image. There exist three problems, i.e., the many-to-many mapping problem between low dynamic range (LDR) images and HDR images, the image quality problem caused by the change of dynamic range and the problem of unpaired LDR–HDR training images. These problems can be solved to some extent through a dual learning framework simultaneously to learn the forward and reverse of camera imaging processes. This procedure is divided into a primary module, to reconstruct HDR from LDR, and a secondary module to reversely mapping the HDR to LDR. The secondary module guides the learning of primary module by constraining the outputs of the primary module. After that, the attention mechanism is used to solve the problem of unnatural perception caused by the change of dynamic range. In the end, with the advantage of our dual learning framework, unpaired data is further explored to train our model, which enriches the training samples. Compared with the state-of-the-art methods, a large number of quantitative and qualitative experiments confirm that our method can achieve better performance.

1. Introduction

In a natural scene, the range of brightness is usually very broad with large light and shade span, showing a lot of visual details. At present, due to the limitation of existing (imperfect) hardware devices, people can only get photos with a certain range of brightness, i.e., common LDR images. Although LDR images to some extent reflect most information of the real scene, there are still some missing details, affecting the overall experience. In order to solve this problem, HDR related technologies are proposed. Compared with a LDR image, a HDR image contains a larger range of brightness, and thus is closer to the real scene. If a LDR image can be reconstructed into a HDR image, it can not only obtain more real and rich image information, but also provides better user experience (Mantiuk et al., 2015).

The key problem of HDR reconstruction from LDR is how to recover the lost brightness and detailed information in the under- and over-exposed areas, respectively (Liu et al., 2020). Generally, there are two ways to reconstruct HDR, i.e., multi-exposure and single-image based approaches. The multi-exposure based approach uses multiple exposure images of the same scene to reconstruct the corresponding HDR image. It usually suffers from the following two problems: firstly, there is a ghost problem (Khan et al., 2006; Mangiat and Stephen, 2010; Srikantha and Sidibé, 2012) caused by incorrect alignment; secondly, multiple exposure images are needed which are not easily satisfied in training data and application scenes. On the contrary, the single-image based approach constructs a HDR image from a single exposure image, which would bring great benefits to practical application but also pose a great challenge to recover the missing brightness value. In this paper, we investigate the single-image based approach. At present, the mainstream methods are to construct deep convolution networks for end-to-end reconstruction of HDR images (Li and Fang, 2019; Marnerides et al., 2018; Eilertsen et al., 2017), or use GAN network (Kim et al., 2020), or cascade multiple secondary networks to reconstruct HDR images (Liu et al., 2020).

Although the above mentioned methods have achieved good results, the following problems still exist. First, the recovery of the lost over-exposed area or under-exposed area is a many-to-many mapping problem due to the different illumination mapping functions of LDR and HDR, resulting in a large solution space. How to reduce the solution space of mapping functions is an important problem for improving the quality of HDR reconstruction. Secondly, in the process of reconstructing HDR, the change of dynamic range brings the change of contrast, saturation and so on, leading to an unnatural appearance of the whole image. How to adjust this change of brightness in HDR to get a better viewing experience is also an important problem. Thirdly, there are not

* Corresponding author.
E-mail address: cvlab.uestc@gmail.com (M. Ye).

https://doi.org/10.1016/j.engappai.2023.105947
Received 4 May 2021; Received in revised form 14 July 2022; Accepted 31 January 2023
Available online 9 February 2023
0952-1976/

many paired training data available. The mainstream methods rely on paired datasets for training, or to simulate the generation of LDR to build paired datasets through known camera characteristics. However, currently most of the images are only LDR images and have lost camera information after shooting. Therefore, a new method enabling the use of these existing large amount of unpaired data is greatly desired to enhance the generalization ability of the model.

To address the above problems, in this paper, we propose a novel **Du**al learning method to reconstruct a **HDR** image, named DuHDR, which is a closed-loop method by learning the forward and reverse camera imaging process to generate HDR images and LDR images. For the many-to-many mapping problem, we use a dual network to increase constraints, reduce solution space and improve learning performance. Ideally, if the mapping process from LDR image to HDR image is optimal, the reconstructed HDR image can obtain the same LDR image as the input through the camera imaging process. Therefore, we can limit the solution from LDR image to HDR image through relatively fixed camera imaging process, so as to reduce the solution space. On the other hand, for improving the image perceptual quality, since different channels and different locations are affected by the change of dynamic range differently, the attention mechanism is used to fit these situations adaptively. For the unpaired image problem, our network itself is a closed-loop dual network, thus can directly use LDR images to further optimize the reconstruction model.

Our contributions can be summarized as follows:

(1) We propose a dual learning framework to the single-image HDR reconstruction task. An extra constraint is introduced by adding the mapping of HDR images to LDR images, helping HDR reconstruction mapping to obtain higher quality HDR images.

(2) We introduce an attention mechanism to adjust the image content level of the HDR reconstructed image, which complements the sensory issues ignored in the reconstructed HDR task, and obtains a balanced and natural image perception.

(3) With our proposed dual learning framework, we pioneer semi-supervised training in the field of HDR image reconstruction. This training and learning method can make full use of the existing unpaired LDR images to increase the diversity of datasets in HDR reconstruction tasks, and improve the generalization ability of the model.

Furthermore, our method has the following practical application scenarios. (1) Through our method, LDR images can be enhanced and HDR images with higher visual quality can be obtained, which can not only make ordinary users shoot high-quality images without professional equipment and professional photography technology, but also improve the display effect of LDR images on HDR displays. (2) When an LDR display needs to show an HDR image, our method can map the HDR image to an LDR image, thereby improving the compatibility of the LDR display. (3) The coding party can convert HDR video into LDR video through the secondary module in our method, and use the existing encoder to directly encode LDR video. The decoding party uses the existing decoder to decode the LDR video, and then restores the HDR video through the primary module in our method.

2. Related works

At present, the mainstream methods of reconstructing HDR images are: multi-exposure images and single-image based approaches.

Multi-exposure images HDR reconstruction method. This technical solution is more general, by processing multiple exposure images of the same scene to predict the missing dynamic range information, so as to realize the reconstruction of HDR (Debevec and Malik, 1997). This approach of multi-exposure images reconstruction HDR, because it has rich dynamic range information, can better predict the larger dynamic range information, but at the same time, if there is any movement of the object in the multi-exposure images, it will bring ghost problem in the reconstructed HDR image. Aiming at the problem of moving ghost in multi-exposure images, Khan et al. (2006), Mangiat and Stephen (2010) and Srikantha and Sidibé (2012) proposed image alignment and post-processing methods to solve this problem. Recently, more effective alignment methods are proposed (Yan et al., 2019; Shangzhe et al., 2018; Kalantari and Ramamoorthi, 2017; Chen et al., 2021b; Ye et al., 2021). At the same time, the detailed information of the over-exposed area is further analyzed and restored (Chen et al., 2021b). Considering the complexity of multi-exposure images approach, single exposure image is commonly used in practical scenes.

Single-image HDR reconstruction method. Single-image based HDR reconstruction can avoid the ghost problem, but it is a great challenge to restore the actual illumination of the scene by only single exposure image. Aiming at the expansion of the brightness range, previous works built models to fit the non-linear process (Francesco Banterle et al., 2009; Akyüz et al., 2007; Banterle et al., 2008, 2006). These methods have high efficiency, but at the same time, due to the rich real scenes, the application scope of these models is relatively small. With the development of deep learning, there are many HDR reconstruction methods based on deep learning. It can be divided into three categories, i.e., end-to-end, multi-exposure simulation, divide and conquer.

The first kind of methods directly reconstruct HDR image from LDR image by CNN, and the process is extremely complex because of the small amount of dynamic range information in a single image (Eilertsen et al., 2017; Zhang and Lalonde, 2017; Marnerides et al., 2018; Yang et al., 2018; Li and Fang, 2019; Yu et al., 2021). This kind of methods has the advantages of fast operation efficiency, less memory consumption and relatively simple process. However, these methods do not well address the problems existing in the HDR image reconstruction process, resulting in poor reconstruction effect and weak generalization ability. The second approach generates multi-exposure images by CNN, and then the HDR image is reconstructed by the enriched dynamic range information (Endo et al., 2017). The effect of this approach is closely related to the quality of the generated multi-exposure images, which is affected by many factors such as over exposure, under exposure or normal exposure, sunlight, light or dim scene. The learning process is still complex. The main advantage of this kind of methods is that it can simulate and restore more exposed scenes, so as to obtain better quality. However, due to the large difference in information between single exposure image and multiple exposure images, the restoration is excessive and produces noise. The final approach learns the inverse process of the camera imaging process to reconstruct HDR image (Liu et al., 2020, 2021; Raipurkar et al., 2021; Chen et al., 2021a). The advantage of this kind of methods is to finely analyze the reconstruction process and carefully design the network to obtain high-quality output. However, only considering the change of brightness in the reconstruction without considering the change of color gamut will lead to unnatural output. In addition, the existing methods are limited by the lack of datasets, resulting in insufficient generalization ability. The proposed method not only comprehensively considers the influencing factors of each link of HDR reconstruction, but also uses dual structure design to expand the training set and solve the problem of insufficient training set.

Dual learning. The dual network contains two mapping functions: $\mathcal{G} : X \rightarrow Y$ and $\mathcal{F} : Y \rightarrow X$, which correspond to primary network and secondary network respectively. During training, the primary network and the secondary network are learned at the same time, and the primary network is constrained by the secondary network, so as to improve the effectiveness of the primary network (Xia et al., 2017; Zhu et al., 2017). It has been widely used in language processing and super-resolution, such as CycleGAN (Zhu et al., 2017), DualGAN (Yi et al., 2017) and DRN (Guo et al., 2020). For the HDR reconstruction task, the main goal is to recover the lost dynamic range information. Because different shooting methods will lead to the loss of dynamic range in varying degrees, resulting in more than one restored image solution, which makes the solution space huge. The dual learning method can add additional constraints to the solution of HDR reconstruction, so as to reduce the solution space. At the same time, the dual learning method can also use the unpaired data set for training, expand the

Table 1
Reviewed technologies.

Technology	Category	References
Multi-exposure image HDR reconstruction	Post-processing	Srikantha and Sidibé (2012)
	Alignment	Khan et al. (2006), Mangiat and Stephen (2010), Yan et al. (2019), Shangzhe et al. (2018), Kalantari and Ramamoorthi (2017), Chen et al. (2021b) and Ye et al. (2021)
	Others	Chen et al. (2021b)
Single-image HDR reconstruction	End-to-end	Eilertsen et al. (2017), Zhang and Lalonde (2017), Marnerides et al. (2018), Yang et al. (2018), Li and Fang (2019) and Yu et al. (2021)
	Multi-exposure simulation	Endo et al. (2017)
	Divide and conquer	Liu et al. (2020, 2021), Raipurkar et al. (2021) and Chen et al. (2021a)
Dual learning	–	Xia et al. (2017), Zhu et al. (2017) and Guo et al. (2020)
Attention mechanism	Channel attention	Cui et al. (2021), Wang et al. (2021) and Li and Fang (2019)
	Spatial attention	Yan et al. (2019)
	Others	Wang et al. (2018) and Lanchantin et al. (2020)
Camera imaging process	–	Debevec and Malik (1997)

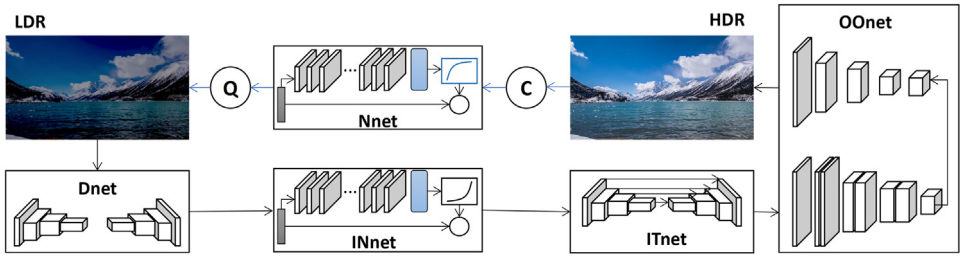


Fig. 1. The overall dual framework (DuHDR). The dual networks consists of three stages according to camera imaging process. Each stage is composed of a primary network and the dual secondary network or an operation. After the dual networks, the overall optimization network (OOnet) is added to further improve the quality.

training set, enrich the training samples and improve the generalization ability of the model.

Attention mechanism. Attention mechanism learns a series of weight parameters through the construction of network, and then use these weight parameters to dynamically weight the image, so as to emphasize the regions we are interested in, and suppress other unrelated regions (Chaudhari et al., 2021). In the field of computer vision, attention mechanism has been applied in object classification (Wang et al., 2018; Lanchantin et al., 2020), object detection (Cao et al., 2020) and super-resolution (Cui et al., 2021; Wang et al., 2021),etc. In the aspect of HDR reconstruction, there are few applications at present. AHDRNet (Yan et al., 2019) uses attention to guide the synthesis of multi-exposure images, and HDRNet (Li and Fang, 2019) uses attention to learn the whole image information to improve the ability of global feature extraction. We use the attention mechanism to focus on adjusting the part that is subject to a large change in dynamic range.

Camera imaging process. We refer to the camera imaging process (Debevec and Malik, 1997) in the process of HDR reconstruction. Although RCPNet has fully analyzed the possible shortcomings in each component of the imaging process, it does not take into account the difficult problem caused by the irreversibility of information loss. In our work, we use a dual network of two-way learning to reduce the difficulty of solution and achieve the goal of improving performance.

The reviewed techniques are summarized in Table 1.

3. The proposed method

As shown in Fig. 1, we use the idea of dual learning to constrain the solution space in order to improve the learning ability, and the attention mechanism to adjust the overall image quality.

In the camera imaging process, the camera first uses a Truncation Process (TP) to obtain the dynamic range value as $\tilde{I}_{clip} = C(H)$, where H represents the real scene with high dynamic range, C represents the truncation function and $\tilde{I}_{clip}$ represents the truncated output. Then the non-linear mapping process is learned by a neural network ($Nnet$) to get $\tilde{I}_n = Nnet(\tilde{I}_{clip})$. In the end, the quantization operation is applied to get the LDR image $\tilde{L} = Q(\tilde{I}_n)$, where Q represents the quantization function. This process can be written as

$$\tilde{L} = Q(Nnet((C(H)))). \tag{1}$$

For the reverse mapping from LDR image to HDR image, a Dequantization network(Dnet) is first learned to recover $\hat{I}_n = Dnet(L)$ where L is a LDR image. Then, an Inverse Non-linear mapping networks (INnet) corresponding to the non-linear mapping process in the forward conversion is developed to recover $\hat{I}_{clip} = INnet(\hat{I}_n)$. Finally, an Inverse Truncation networks ($ITnet$) is learned to recover HDR image $\hat{H} = ITnet(\hat{I}_{clip})$. This reverse mapping process can be summarized as

$$\hat{H} = ITnet(INnet(Dnet(L))). \tag{2}$$

We call the module on the path from LDR to HDR as the corresponding primary module and the module on the path from HDR to LDR as secondary module.

Finally, an overall optimization network (OOnet) is designed to improve the subjective perceived quality of reconstructed HDR and obtain the final output $\hat{H}_{OO} = OOnet(\hat{H})$.

On the paired dataset, firstly, the primary and secondary modules at each stage are trained independently to obtain an initial model by individual and joint training. Then the OOnet and whole network are trained. After that, the unpaired dataset is used to fine-tune the primary and secondary modules. Compared with the supervised training approach using only paired dataset, the proposed semi-supervised training using unpaired dataset can obtain higher quality output images. In the following, the proposed dual neural networks in three stages and the overall optimization network are explained in detail.

3.1. Dequantization network (Dnet)

In the camera imaging process, the 32-bit image is quantized to an 8-bit LDR image. This quantization operation is mainly carried out

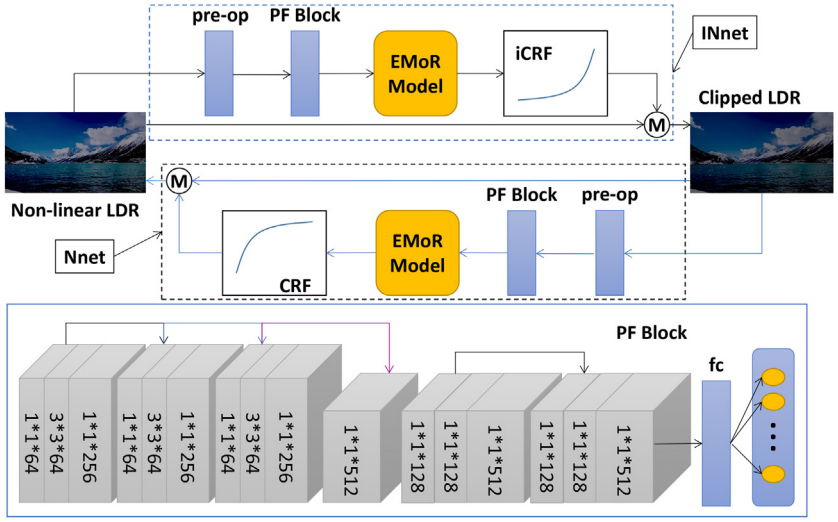


Fig. 2. Dual non-linear mapping learning. The Nnet learns CRF curve to map the clipped LDR image to non-linear LDR image while the INnet learning ICRF curve to inversely map non-linear LDR image to the clipped LDR image.

through the following formula,

$$\tilde{L} = Q(\tilde{I}_n) = \lfloor 255 \times \tilde{I}_n + 0.5 \rfloor / 255. \tag{3}$$

where $\tilde{L}$ is the quantization result of $\tilde{I}_n$. Dnet and the quantization function forms dual learning. The task of the Dnet is to recover the lost information and suppress the noise, while the quantization function constrains the results of the Dnet by making sure no new noise is generated. The Dnet uses a U-Net network structure with 6 stacks of layers, each with two convolutional layers.

Training on paired dataset: In the training process of Dnet, we take L as the input of Dnet and bring the output of Nnet $\tilde{I}_n$ and an inverse quantized image $\hat{I}_n$ generated from Dnet into the loss function for learning. For individual training, the loss function is

$$loss_{dnd} = \|\hat{I}_n - \tilde{I}_n\|_2^2. \tag{4}$$

For joint dual training process of Dnet, it takes the quantization operation into the training of Dnet. The loss function is

$$loss_{dn} = \|\hat{I}_n - \tilde{I}_n\|_2^2 + \lambda_{dn\_dual} \cdot \|\tilde{L} - L\|_2^2 \tag{5}$$

where λ_{dn_dual} is the dual weight parameter.

Training on unpaired dataset: The unpaired dataset, only containing the LDR images L, is used as the input of Dnet and the output $\hat{I}_n$ of Dnet as the input of quantization operation. The loss function is

$$loss_{dnu} = \|Q(Dnet(L)) - L\|_2^2. \tag{6}$$

3.2. Dual non-linear mapping

The non-linear mapping process maps the brightness value obtained by the camera to the brightness value of the digital image through a CRF (camera response function) and different functions may be used in different cameras. Since the brightness change perceived by the human eye is not a linear relationship with the real brightness value change, the CRF curve can map the brightness value obtained by the camera to the brightness value observed by the human eye. So a method of estimating the CRF function is needed if we want to re-formulate the camera imaging process. Similarly, to recover a HDR image, we also need the inverse CRF curve (ICRF). Therefore, as shown in Fig. 2, the dual learning of the non-linear mapping process is further divided into two networks: Nnet and INnet. The task of the INnet is to learn the ICRF process while Nnet learns CRF which guides the learning of INnet.

A CRF or ICRF curve is usually discretized by uniformly sampling 1024 points between [0, 1]. According to the EMoR model (Grossberg and Nayar, 2003), a CRF or ICRF curve can be represented by k basis vectors, thus our problem is transformed into solving the coefficients of k basis vectors ($k = 11$ in our work). Accordingly, the goal of Nnet is to find the coefficient vector c^{crf} of CRF and the goal of INnet is to find the coefficient vector c^{icrf} of ICRF, respectively.

Since the CRF curve is an one-to-one non-linear monotonic curve, the ICRF curve has the same theoretical space as CRF curve. Therefore, the Nnet and INnet can use a same network structure. Moreover, considering the edge and histogram are helpful to estimate CRF or ICRF curves (Lin et al., 2004; Lin and Zhang, 2005), edge information obtained by Sobel filter and the image histogram are used as input. We use ResNet-18 (Li and Peers, 2017; He et al., 2016) as the main body to obtain the coefficient vector,

$$c = fc(Res(pre\text{-}op(x))), \tag{7}$$

where the function $pre\text{-}op(\cdot)$ means the preprocessing of the image x, $Res(\cdot)$ is ResNet-18 and $fc(\cdot)$ means the fully connected layer.

According to the EMoR model, with the coefficient c^{icrf}, the ICRF curve can be obtained by using the following Eq. (8):

$$g^{icrf} = g_0 + H c^{icrf\,T} \tag{8}$$

where c^{icrf} is the coefficient and g_0 is the mean curve. $H = [h_1, h_2, \ldots, h_k]$ and h_i is the basis vector in 1024-dimensional space. g^{crf} can be obtained in the same way by using c^{crf}. Then for INnet, the process of mapping an image with an ICRF curve can be represented as the following,

$$\hat{I}_{clip} = \mathcal{M}(\hat{g}^{icrf}, \hat{I}_n), \tag{9}$$

and the mapping for Nnet is

$$\tilde{I}_n = \mathcal{M}(\tilde{g}^{crf}, x) \tag{10}$$

where $\mathcal{M}$ represents the process of mapping an image using a CRF $\hat{g}^{crf}$ (or ICRF $\hat{g}^{icrf}$) curve based on EMoR model. The input x can be formed in two cases: (1) $\tilde{I}_{clip} = C(H)$ for paired training data; (2) $\hat{I}_{clip}$ for unpaired LDR training data and dual training.

Training on paired dataset: In the individual training process, a non-linear LDR image reconstruction loss for Nnet is defined as

$$loss_{Nnet} = \|\tilde{I}_n - \hat{I}_n\|_2^2. \tag{11}$$

For training INnet, the loss is

$$loss_{INnet} = \|\hat{I}_{clip} - \tilde{I}_{clip}\|_2^2. \tag{12}$$

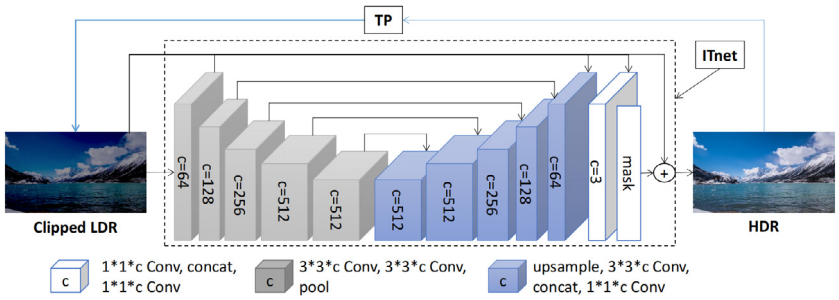


Fig. 3. From HDR image to the clipped LDR image, the truncation process (TP) is applied. The Inverse Truncation network (ITnet) contains a 6-layer auto-encoder and an over-exposed mask employed to reconstruct HDR image.

In the joint training process, the input of joint training comes from the output of pre-trained Dnet, and the output image of INnet is used as the input of Nnet, and the loss is

$$loss_{dnm} = loss_{INnet} + \lambda_{dnm\_dual} \cdot loss_{Nnet} + \lambda_{dnm\_dn} \cdot loss_{dn} \tag{13}$$

where λ_{dnm_dual} and λ_{dnm_dn} are the balance parameters respect to the Nnet loss and quantization loss $loss_{dn}$ in Eq. (5).

Training on unpaired dataset: The input and output of the training process are the same for joint training on paired dataset, and the loss is

$$loss_{dnmu} = \|Nnet(INet(\hat{I}_n)) - \hat{I}_n\|_2^2. \tag{14}$$

3.3. Inverse Truncation network (ITnet)

This module forms dual learning by two parts: Truncation Process (TP) and ITnet as shown in Fig. 3. The truncation operation follows the method in He et al. (2013). To recover the over-exposed area, our ITnet uses an auto-encoder network with skip connections(SAE) (Eilertsen et al., 2017) and an over-exposed mask, $mask = max(0, x_{clip} - \alpha)/(1-\alpha)$, i.e.,

$$\hat{H} = ITnet(x_{clip}) = x_{clip} + mask \cdot SAE(x_{clip}), \tag{15}$$

where x_{clip} also has two cases of input: $\tilde{I}_{clip}$ for paired training data and $\hat{I}_{clip}$ for unpaired LDR training data. In this paper, we set $\alpha = 0.95$.

The loss of the ITnet is formulated by three parts: log domain loss, visual perception loss (Johnson et al., 2016) and normal area loss. The log domain loss is $loss_{log} = \|\log(\hat{H}) - \log(H)\|_2^2$ to avoid large pixel value; the visual perception loss is $loss_{prec} = \|(vgg16(\hat{H}) - vgg16(H)) \cdot mask\|_2^2$ which is mainly responsible for recovering the lost information in the over-exposed area. In order to restrain the effect of this module on other normal area to avoid introducing extra noise, the normal area loss is $loss_{nor} = \|(\tilde{I}_{clip} - I_{clip}) \cdot (1 - mask)\|_2^2$, where $I_{clip} = min(1, H)$. The last term will be used in the joint training.

Training on paired dataset: For the individual training process, the output $\hat{H}$ of ITnet network is brought into the loss function for learning. The loss function is

$$loss_{itnd} = loss_{log} + \lambda_{prec} \cdot loss_{prec}. \tag{16}$$

For the joint training process, the input is $\tilde{I}_{clip}$, and the loss function is

$$loss_{itn} = loss_{log} + \lambda_{prec} \cdot loss_{prec} + \lambda_{itn\_dual} \cdot loss_{nor} \tag{17}$$

where λ_{prec} and λ_{itn_dual} are balanced parameters. In this training phase, the paired dataset and unpaired dataset are used in an iterative way, as shown in the following,

$$\tilde{I}_{clip_i} = C(\hat{H}_i), \tag{18}$$

$$\hat{H}_{i+1} = ITnet(\tilde{I}_{clip_i}) \tag{19}$$

where $\hat{H}_i$ is the ith output result of the ITnet; $\tilde{I}_{clip_i}$ represents the ith output of our truncation operation; and $\hat{H}_{i+1}$ is the $(i+1)$th output result of the ITnet. In this way, we can constantly use the recovered details to infer and predict more details, thus reducing the solution space of the recovery problem and improving the learning ability. With the increase of the number of iterations, the solution space tends to be decreased, and the details tend to increase, but it is inclined to bring over fitting problems, and servely increase the amount of calculation. In this paper, a total of three iterations are used. **Training on unpaired dataset:** The input is the output $\hat{I}_{clip}$ of pre-trained INnet, and the loss function used is

$$loss_{itnu} = \|C(ITnet(\hat{I}_{clip})) - \hat{I}_{clip}\|_2^2. \tag{20}$$

Similar to the joint training process on paired dataset, this training phase is also trained in an iterative way.

3.4. Overall Optimization network (OOnet)

By the above dual learning process, there still exist some problems in the final $\hat{H}$ image. The first problem is the unnatural contrast problem. Due to the different restoration degree of different regions, the over-exposed area of $\hat{H}$ looks particularly bright, while other area of $\hat{H}$ looks relatively dark. The second problem is the unnatural perception problem after dynamic range expansion. In addition to the above two main problems, there are still some problems such as noise.

Therefore, OOnet is developed to deal with the abovementioned problems for $\hat{H}$ as

$$\hat{H}_{OO} = OOnet([\hat{H}, \hat{I}_{clip}, \hat{I}_n]). \tag{21}$$

OOnet is constructed based on the U-net and consists of three parts, i.e., the encoder part, adjustment part and reconstruction part. The OOnet structure is shown in Fig. 4.

In the encoder part, the DownBlock is used to extract key feature information and filter out irrelevant information. In this paper, pooling and RFB nets are used to form a DownBlock as shown in Fig. 4. RFB network (Liu et al., 2018) is used, which is designed according to the receptive field properties of human visual system is used to solve the unnatural contrast problem. The RFB extracts features of three scales and fuse them with input. The fusion coefficient is 0.1. The DownBlock adds batch normalization to accelerate convergence and improve performance (Pan et al., 2020),

$$y_i^{\downarrow} = Relu(DownBlock(x_i)) \tag{22}$$

where $y_i^{\downarrow}$ represents the features obtained by the ith down sampling process, and x_i represents the input of the ith DownBlock.

The unnatural perception problem is because different channels and different locations are affected by the change of dynamic range. Therefore, it is necessary to adjust the weight relationship between feature maps to make the image close to the real scene. Secondly, since the change of dynamic range affects the saturation of the image and it is related to the spatial positional relationship between pixels, it is necessary to consider the spatial influencing factors. A multi-scale channel learning mechanism (MS-SE Block) is developed, as shown

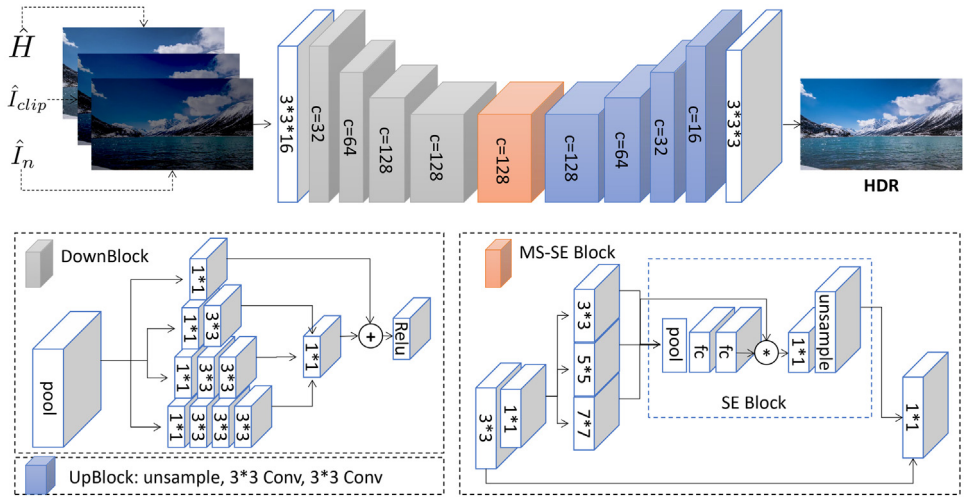


Fig. 4. The structure of OOnet. The encoder part consists of a convolution layer and four DownBlocks. Then, the features are adjusted by a MS-SE Block. Finally, the image is reconstructed by a reconstruction part which consists of 4 UpBlocks and a convolution layer.

in Fig. 4. MS-SE Block consists of a channel attention module (SE Block) (Hu et al., 2020) to adjust the weight relationship between features, and a multi-scale feature extraction module to solve the spatial influence. The final feature extraction results y_{MS_SE} can be written as

$$y_{MS\_SE} = Relu(MS\_SE(y_4^{\downarrow})). \tag{23}$$

The reconstruction part consists of four UpBlocks and a convolution layer. The loss of OOnet is defined as follows,

$$loss_{OO} = \|\hat{H}_{OO} - H\|_2^2. \tag{24}$$

Training on paired dataset: For OOnet, there is only the joint training process. By combining Dnet, INnet and ITnet, the input of OOnet is $\hat{H} = ITnet(INnet(Dnet(L)))$. The HDR image H corresponding to the LDR image L is taken as the target and brought into the loss function $loss_{OO}$ to adjust the OOnet.

Training on unpaired dataset: The unpaired LDR image L is used as the following: First, $\hat{H}_{unpair} = OOnet(ITnet(INnet(Dnet(L))))$; then, the loss is

$$loss_{OOu} = \|\tilde{L}_{unpair} - L\|_2^2 \tag{25}$$

where $\tilde{L}_{unpair} = Q(Nnet(C(\hat{H}_{unpair})))$.

Remark. According to the theoretical analysis of dual regression scheme (Guo et al., 2020), it can be similarly shown that the dual task helps the primary network obtain a smaller generalization bound, reduce the solution space and improve the prediction quality. For the HDR reconstruction task, let $L_P(P(x), y) + \lambda L_S(S(P(x)), x)$, $x \in \mathcal{X}, y \in \mathcal{Y}$ be a mapping from $\mathcal{X} \times \mathcal{Y}$ to $[0, C]$, where $x \in \mathcal{X}$ is the LDR image, $y \in \mathcal{Y}$ is the HDR image; P and $L_P(\cdot)$ are the mapping from LDR images to HDR images and its loss function respectively; S and $L_S(\cdot)$ are the mapping from HDR images to LDR images and its loss function respectively. Since the dual regression problem is essentially consistent with the mapping between HDR and LDR, according to Theorem 3 in Guo et al. (2020), we can get $\Phi(P, S) \leq \Phi(P)$, where $\Phi(P, S)$ is the generalization bound of dual task scheme, and $\Phi(P)$ the generalization bound of standard supervised scheme. So the generalization bound of the dual learning is smaller than that of the standard supervised learning, which means that the dual task reduces the solution space of the main task and helps the main task obtain higher solution quality.

Algorithm 1 Training process of DuHDR algorithm based on paired data

Input: Paired data D_p

```
# Randomly initialize DuHDR networks;
while not convergent do
    Sample data L, H from D_p;
    Update Dnet by minimizing loss_dnd in Eq. (4)(individual training); Then update Dnet by minimizing loss_dn in Eq. (5) (joint training);
    Update Nnet and INnet by minimizing loss_Nnet and loss_INnet in Eqs. (11)-(12) (individual training); Then update Nnet and INnet by minimizing loss_dnm in Eq. (13) (joint training);
    Update ITnet by minimizing loss_itnd in Eq. (16)(individual training); Then update ITnet by minimizing loss_itn in Eq. (17) (joint training);
end while
while not convergent do
    Update OOnet by minimizing loss_OO in Eq. (24);
end while
Fine-tune the whole networks based on loss_OO.
```

Algorithm 2 Training process of DuHDR algorithm based on unpaired data

Input: Unpaired data D_u, pre-trained DuHDR model;

```
while not convergent do
    Sample data L from D_u;
    Update Dnet by minimizing loss_dnu in Eq. (6);
    Update Nnet and INnet by minimizing loss_dnmu in Eq. (14);
    Update ITnet by minimizing loss_itnu in Eq. (20);
end while
while not convergent do
    Update OOnet by minimizing loss_OOu in Eq. (25);
end while
```

3.5. Algorithm summarization

Through the detailed introduction of the above sub-modules, this section summarizes our algorithm as follows. For the paired data set D_p, the dual neural networks at each stage are trained independently. Then, the Overall Optimization Network (OOnet) is trained.

Table 2
Comparison results with the state-of-the-art HDR reconstruction methods. Except the model size (**SIZE**) and time complexity (**FLOPs**), all scores in the table are HDR-VDP average.

Method	Training datasets	HDR-Real	HDR-Eye	RAISE	SIZE	FLOPs
HDRCNN (Eilertsen et al., 2017)	HDR-Synth, HDR-Real	8.16	6.41	9.27	29.44M	242.81G
DrTMO (Endo et al., 2017)	HDR-Synth, HDR-Real	8.23	6.67	9.43	348.99M	3.12T
ExpandNet (Marnerides et al., 2018)	HDR-Synth, HDR-Real	8.05	6.49	9.31	0.46M	53.4G
RCPNet (Liu et al., 2020)	Pre-trained model (Liu et al., 2020)	8.32	6.69	9.51	29.13M	451.04G
Ours(DuHDR)	HDR-Synth, HDR-Real	8.37	6.73	9.54	30.81M	452.43G
Ours*(DuHDR)	HDR-Synth, HDR-Real, Unpair	8.36	6.75	9.55	30.81M	452.43G

For the unpaired dataset D_u, it can be used to fine-tune the pre-trained model by joint training. All the process is shown in Algorithms 1 and 2. For more details, please refer to the source code at github.com/DoomsdayLeaf/DualHDR.git.

4. Experiments

In this section, we will compare the state-of-the-art HDR reconstruction methods on public datasets, by using standard evaluation criterions. Then, the parameters involved in this work and each network will be analyzed to illustrate their effects.

4.1. Dataset and evaluation criteria

Dataset. We used four datasets: HDR-Synth (Liu et al., 2020), HDR-Real (Liu et al., 2020), HDR-Eye (Nemoto et al., 2015), RAISE (Dang-Nguyen et al., 2015). Among them, HDR-Eye and RAISE are two common datasets, which contain all kinds of scenarios and are widely used in various HDR reconstruction methods. HDR-Synth and HDR-Real are datasets produced in Liu et al. (2020), which contain a large number of scenes and known camera information. In our work, we use HDR-Synth and HDR-Real as training sets, HDR-Real, HDR-Eye and RAISE as test datasets.

Evaluation. HDR images are very different from LDR images in both value gamut and color gamut, so traditional evaluation indicators cannot better evaluate the quality of HDR images. HDR-VDP builds an image evaluation algorithm according to the characteristics of HDR images. In recent years, many excellent works (Liu et al., 2020, 2021; Raipurkar et al., 2021; Yu et al., 2021) in the field of HDR image reconstruction have adopted this indicator as a quantitative evaluation standard.

We use the latest HDR-VDP-3 (Mantiuk et al., 2011) to evaluate the quality of reconstructed HDR. HDR-VDP is a visual measure used to compare the reference image and the test image, and provides prediction information in terms of visibility and quality. The principle is to use the algorithm to simulate the process of human eyes viewing images, and predict the different points of two different images viewed by human eyes, thus obtain the quality of the predicted HDR image. The computation process of HDR-VDP-3 is complex, for detail, please refer to the paper (Mantiuk et al., 2011). We directly use the source code offered in Mantiuk et al. (2011).

Before calculating the quality value of the predicted HDR image by using the HDR-VDP source code, we need to normalize the predicted HDR and the reference real HDR images by the method in Marnerides et al. (2018), and map their ranges to the specified display brightness range as follows,

$$y_{i,j} = \frac{L_{max} - L_{min}}{x_{max} - x_{min}} \cdot (x_{i,j} - x_{min} + \epsilon) + x_{min} \tag{26}$$

where $x_{i,j}$ represents the pixel value of the input at (i, j), $y_{i,j}$ represents the pixel value of the output at (i, j), x_{min} and x_{max} respectively represent the minimum and maximum values of the input. L_{min} and L_{max} respectively represent the minimum and maximum brightness values of the display. The parameter ϵ is set as 10^{-5}.

4.2. Comparison results

We compare four representative methods: divide and conquer method (RCPNet) (Liu et al., 2020), end-to-end method (HDRCNN) (Eilertsen et al., 2017), simulated multi-exposure method (DrTMO) (Endo et al., 2017), multi-feature end-to-end method (ExpandNet) (Marnerides et al., 2018). Since the training dataset of RCPNet is consistent with ours, so the pre-trained model is directly used. HDRCNN, ExpandNet and DrTMO are retrained by the datasets HDR-Synth and HDR-Real with the provided models.

Quantitative comparison. Table 2 shows the average scores of each method tested in HDR-Real, HDR-Eye and RAISE datasets. All methods were trained on HDR-Synth and HDR-Real datasets, while Ours*(DuHDR) represents the results of further trained model on the unpaired dataset. Compared with the state-of-the-art methods, our method has achieved better performance on all four datasets. The results show that our method (DuHDR) has improved by 0.04 on HDR-Real, 0.06 on HDR-Eye and 0.04 on RAISE. From the quantization results, the gain is not very obvious. It is because that the gain is approaching an upper limit with the deepening of the current researches.

Visual comparison. Three kinds of images are used for comparison. As shown in Fig. 5, the image (**a**) is a single light source indoor scene with strong light and dark pairs and rich color types. The image (**b**) is the scene of insufficient exposure of near objects caused by sunlight. The image (**c**) is the outdoor scene of multi-light buildings at night. The selected images are from the test set, and in the following experiments, our model adopts the final model further trained by unpaired data. From the experimental results shown in Fig. 5, we can obtain the following observations. The HDRCNN method and ExpandNet method are not enough to restore the details in the over-exposed area; while the DrTMO method restores excessively and produces obvious noise, and the contrast is reduced; and the restoration effect of RCPNet is insufficient saturation. Our method balances the contrast and saturation while restoring the information, and improves the visual quality.

By analyzing Fig. 5(2), the HDRCNN method can recover some details of the under-exposed area, but it can be seen from Fig. 5(2)(a2) that there are still some areas that have not been recovered, and there are certain color differences in various colors. It is not difficult to see from Fig. 5(2)(c) that the recovery effect is not good for the overall bright environment, and the detail recovery is not enough.

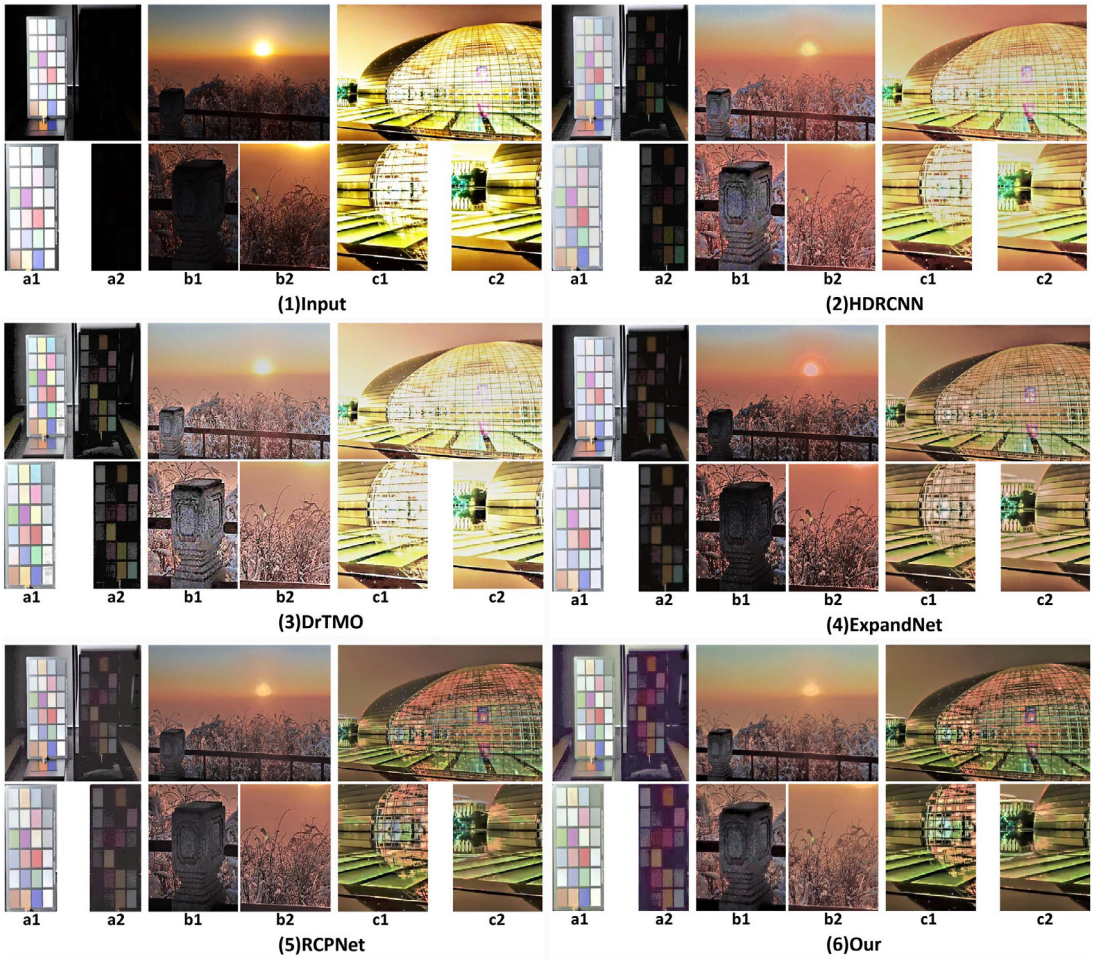


Fig. 5. Visual comparison. Because HDR images cannot be displayed directly, we use a unified tone mapping method (Liang et al., 2018) to show the images.

For ExpandNet, by analyzing Fig. 5(4)(a2), (4)(b1) and (4)(b2), it is insufficient in detail recovery of under-exposed area, and there is color difference in some colors, while Fig. 5(4)(c) shows that there is a certain degree of recovery for details of over-exposed area.

Through Fig. 5(3)(a2), (3)(b1) and (3)(b2), it can be seen that DrTMO can recover the details of under-exposed area, but it seems that there are relatively more noises, and the overall contrast of the image is small and there is fading. Moreover, it can be seen from Fig. 5(3)(a1) that artifacts are produced in the over-exposed area. From Fig. 5(3)(a), it can be seen that the detail recovery of the over-exposed area is not very good.

From Fig. 5(5)(a2), (5)(b1) and (5)(b2), it shows that RCPNet can restore the details of under-exposed area. However, it can be seen from Fig. 5(5)(a2) that the image saturation is not enough and the color block has color difference. It can be seen from Fig. 5(5)(c) that the details of the over-exposed area are restored to a certain extent, but artifacts appear in the bright block area in Fig. 5(5)(c1).

On the whole, HDRCNN method is not enough in detail restoration for both over-exposed area and under-exposed area; ExpandNet method is insufficient in detail restoration of under-exposed area; DrTMO method is excessive on over-exposed area, artifacts appear, and the overall contrast of the image is low; RCPNet method is good in detail restoration of both over-exposed area and under-exposed area, but the image saturation is not enough and there are chromatic aberration and some artifacts in the over-exposed area. Compared with these methods, our method is better in detail restoration of over-exposed area or under-exposed area, and the overall contrast and saturation of the image are well adjusted, and good appearance quality is achieved.

Complexity analysis. As most of deep neural network based methods, time complexity is evaluated using FLoating point OPerations (FLOPs) (Molchanov et al., 2017) and space complexity is measured by the size of model parameters. As shown in Table 2, our method has a certain increase in computational complexity compared to the state-of-the-art methods, but our method achieves better quality. For the four parts of our method: DNet, INnet, ITnet and OOnet, the corresponding FLOPs are 37.94G, 54.50G, 320.52G and 39.47G corresponding to the four stage and the total FLOPs is 452.43G. Furthermore, we also run our inference codes on three datasets: HDR-Real, HDR-Eye and RAISE based on an NVIDIA 2080 GPU. The average time is 0.9 s, 0.6 s, 2.2 s and 1.0 s corresponding to the four stages and the total average time is 4.7 s.

From the above experiments, it can be observed that our method achieves better performance than the state-of-the-art methods in both quantitative results and visual analysis results. At the same time, it can be seen from the quantitative comparison experiments that using unpaired datasets can improve the performance. However, from the complexity analysis part, our method is not real-time. The major cost is ITnet, which occupies 46.8% of the total time-consuming. We think this cost is deserved. It can be seen from the visual perception experiments that the better quality images are generated, which are excellent in detail recovery in each region. With the help of feature adjustment, our generated images have a more natural visual experience.

4.3. Parameter discussion

The parameters are divided three groups according to Dnet, Dual Nonlinear Mapping and ITnet. The parameter analysis experiments

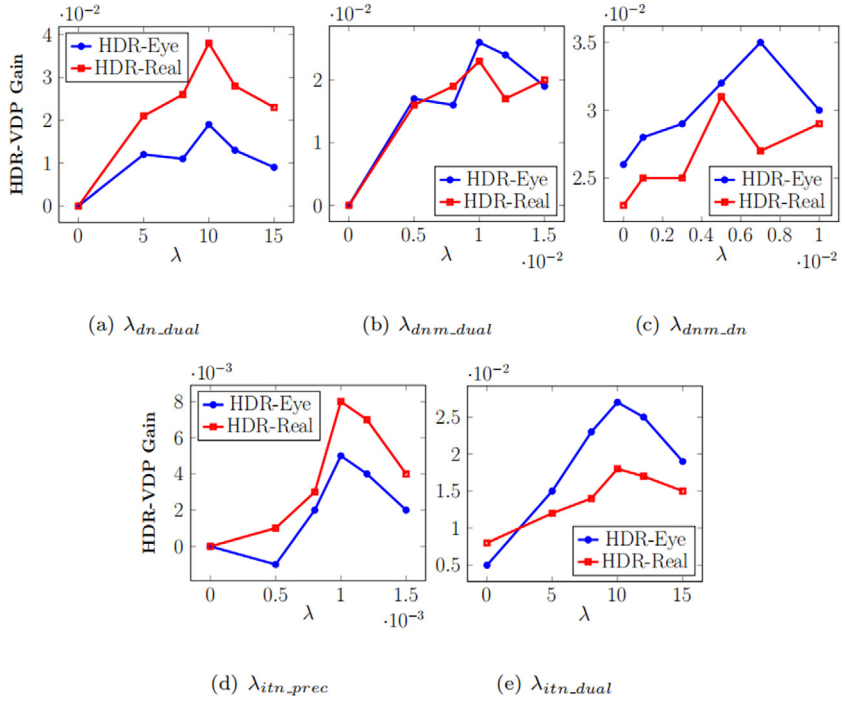


Fig. 6. Parameter analysis. The performance of parameters λ_{dn_dual}, λ_{dnm_dual}, λ_{dnm_dn}, λ_{itn_prec} and λ_{itn_dual} are shown respectively. From all our experiments, the setting of parameters has a relatively stable range.

are performed on the datasets HDR-Real and HDR-Eye. During experiments, we fix other parameters and analyze how the change of a parameter value affects the performance of our model.

The parameter λ_{dn_dual} for training Dnet. As shown in Eq. (5), λ_{dn_dual} is the dual weight coefficient on inverse quantization operation which can filter construction noise. So its value is usually larger. Through the experimental observation of Fig. 6(a), we found that when the value is about 10, the performance is best.

The parameters λ_{dnm_dn} and λ_{dnm_dual} for training Dual Non-linear Mapping. As shown in Eq. (13), λ_{dnm_dn} represents the degree of influence of Dnet loss on Dual Non-linear Mapping. It does not effect the training of Dual Non-linear Mapping very much, so λ_{dnm_dn} is set to a smaller value. From Fig. 6(c), it can be concluded that when the value of λ_{dnm_dn} is 0.005, the performance is the best. It should be noted that the bad Dnet loss does affect the training of Dual Non-linear Mapping. When λ_{dnm_dn} drops to zero, the performance decreases. λ_{dnm_dual} is the balance parameter between Nnet and INnet. It can be observed that the best value is 0.01 in Fig. 6(b).

The parameters λ_{itn_prec} and λ_{itn_dual} for training ITnet. As shown in Eq. (17), λ_{itn_prec} indicates the impact of visual perceptual loss on training ITnet. Since the training of ITnet mainly comes from $loss_{log}$, and the loss of visual perception not only suppresses the noise, but also suppresses the recovery of dynamic range, so λ_{itn_prec} value should be small. As shown in Fig. 6(d), the best value is 0.001. λ_{itn_dual} is in charge of the weight of loss in normal area on training ITnet. Since the normal area is generally relatively large, which has a great impact on the training of ITnet, so λ_{itn_dual} should be set to a larger value. It can be found in Fig. 6(e) that when λ_{itn_dual} is set to 10, the performance is the best.

4.4. Ablation experiment

Ablation experiments are conducted to verify the effectiveness of our network design and training strategy. Our network is divided into four parts. Since Dnet and ITnet are relatively simple, we do ablation analysis on two networks: Dual Non-linear Mapping and OOnet. All ablation experiments are performed on HDR-Eye dataset.

Table 3
The effectiveness are shown by using different modules to participate in training dual non-linear map. The results obtained by only INnet is considered as the baseline.

INnet	Nnet	Dnet	TP	HDR-VDP
✓	–	–	–	6.672
✓	✓	–	–	6.698
✓	✓	✓	–	6.704
✓	✓	✓	✓	6.712

Table 4
Analysis of the effectiveness of each module of OOnet. The output $\hat{H}$ of ITNet is the baseline.

Baseline	U-Net	RFB	MS-SE	HDR-VDP
✓	–	–	–	6.712
✓	✓	–	–	6.695
✓	✓	✓	–	6.719
✓	✓	✓	✓	6.731

Dual Non-linear Mapping. Here we mainly show the training strategy. The modules involved in training dual non-linear mapping are Dnet, TP, INnet and Nnet. We verify the performance by adding these modules one by one through the control variable method (Liu et al., 2020). The experimental results are shown in Table 3. From the experimental results, by constantly adding the relevant modules, all modules can improve the performance.

OOnet. OOnet network consists of three core sub-networks: U-Net, RFB and MS-SE. As the experimental results shown in Table 4, we find that using U-Net to adjust and optimize cannot improve the image quality. After we added the designed network RFB and MS-SE, it has been improved. The main reason why we still use U-Net is that the training of OOnet needs to combine Dnet, Dual Non-linear Mapping and ITnet, which lead to a large amount of memory. Through U-Net, we can use less memory to complete the overall tuning.

5. Conclusion

In order to solve the problems of unnatural color gamut, large solution space issue of possible mapping functions and insufficient dataset

in the existing single frame HDR image reconstruction methods, this paper proposed a novel method for single-image HDR reconstruction by dual learning camera imaging process. The dual learning method is used to optimize the training of the reconstruction networks. Then the attention mechanism is introduced to solve the appearance quality problem caused by the dynamic changes of image contrast and saturation. Finally, the dual network is further trained by making full use of the unpaired data. As a result, the proposed network is able to solve the above problems such as the large solution space issue of possible mapping functions, chromatic aberration and insufficient dataset, which appear in conventional HDR reconstruction.

More importantly, our proposed method is not only a creative semi-supervised learning method in HDR reconstruction task, where HDR reconstruction task can make full use of unpaired datasets, but also has the following practical application values. First, by using our method, the mutual conversion between LDR image and HDR image can be completed, which can not only improve the quality of LDR image, but also improve monitor compatibility. Second, we can convert HDR video into LDR video through our method for encoding, decoding and transmission, and finally restore HDR video with our network, so as to realize HDR video encoding and decoding. Therefore, our proposed method not only solves the problems existing in the traditional HDR reconstruction methods and obtains higher quality output, but also expands the application scenario of HDR reconstruction task so that it is no longer limited to the application scenario of improving the quality of LDR imagery.

Although our method has achieved good performance, our method cannot be applied to scenes with strict real-time requirements. Through the complexity analysis in the experiment section, the main time-consuming part is the recovery of over-exposed area. A better feature extraction method can reduce the demand of feature maps, so as to reduce the network parameters and computations, and reduce the time complexity. This will be studied in the future.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used in this paper is public shared by other papers which are cited in the manuscript.

Acknowledgments

This work was supported in part by National Key R&D Program of China (2018YFE0203900) and Sichuan Science and Technology Program, PR China (2020YFG0476).

References

Akyüz, A., Fleming, R., Riecke, B., Reinhard, E., Bülthoff, H., 2007. Do HDR displays support LDR content?: A psychophysical evaluation. ACM Trans. Graph. 26, 38.1–38.7.

Banterle, F., Ledda, P., Chalmers, A., 2008. Expanding low dynamic range videos for high dynamic range applications. In: Proceedings of the 24th Spring Conference on Computer Graphics. pp. 33–41.

Banterle, F., Ledda, P., Debattista, K., Chalmers, A., 2006. Inverse tone mapping. In: Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia. pp. 349–356. http://dx.doi.org/10.1145/1174429.1174489.

Cao, J., Chen, Q., Guo, J., Shi, R., 2020. Attention-guided context feature pyramid network for object detection. arXiv:2005.11475.

Chaudhari, S., Mithal, V., Polatkan, G., Ramanath, R., 2021. An attentive survey of attention models. arXiv:1904.02874.

Chen, X., Liu, Y., Zhang, Z., Qiao, Y., Dong, C., 2021a. HDRUNet: Single image HDR reconstruction with denoising and dequantization. arXiv:2105.13084.

Chen, J., Yang, Z., Chan, T.N., Li, H., Hou, J., Chau, L.-P., 2021b. Attention-guided progressive neural texture fusion for high dynamic range image restoration. arXiv:2107.06211.

Cui, Z., Leng, J., Liu, Y., Zhang, T., Quan, P., Zhao, W., 2021. SKNet: Detecting rotated ships as keypoints in optical remote sensing images. IEEE Trans. Geosci. Remote Sens. 1–15. http://dx.doi.org/10.1109/TGRS.2021.3053311.

Dang-Nguyen, D.-T., Pasquini, C., Conotter, V., Boato, G., 2015. RAISE: a raw images dataset for digital image forensics. In: Proceedings of the 6th ACM Multimedia Systems Conference. pp. 219–224.

Debevec, P.E., Malik, J., 1997. Recovering high dynamic range radiance maps from photographs. In: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques. pp. 369–378.

Eilertsen, G., Kronander, J., Denes, G., Mantiuk, R.K., Unger, J., 2017. HDR image reconstruction from a single exposure using deep CNNs. ACM Trans. Graph. 36, 6. http://dx.doi.org/10.1145/3130800.3130816.

Endo, Y., Kanamori, Y., Mitani, J., 2017. Deep reverse tone mapping. ACM Trans. Graph. (Proc. of SIGGRAPH ASIA 2017) 36 (6).

Francesco Banterle, K.D.., Artusi, A., Pattanaik, S., Myszkowski, K., Ledda, P., Chalmers, A., 2009. High dynamic range imaging and low dynamic range expansion for generating HDR content. Comput. Graph. Forum. 28 (8), 2343–2367.

Grossberg, M., Nayar, S., 2003. What is the space of camera response functions? In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings, Vol. 2. pp. II–602. http://dx.doi.org/10.1109/CVPR.2003.1211522.

Guo, Y., Chen, J., Wang, J., Chen, Q., Cao, J., Deng, Z., Xu, Y., Tan, M., 2020. Closed-loop matters: Dual regression networks for single image super-resolution. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5406–5415.

He, K., Sun, J., Tang, X., 2013. Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. 35, 1397–1409. http://dx.doi.org/10.1109/TPAMI.2012.213.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. http://dx.doi.org/10.1109/CVPR.2016.90.

Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E., 2020. Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. 42 (8), 2011–2023.

Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision, Vol. 9906. pp. 694–711. http://dx.doi.org/10.1007/978-3-319-46475-6_43.

Kalantari, N.K., Ramamoorthi, R., 2017. Deep high dynamic range imaging of dynamic scenes. ACM Trans. Graph. (Proceedings of SIGGRAPH 2017) 36.

Khan, E.A., Akyuz, A.O., Reinhard, E., 2006. Ghost removal in high dynamic range images. In: 2006 International Conference on Image Processing. pp. 2005–2008. http://dx.doi.org/10.1109/ICIP.2006.312892.

Kim, S.Y., Oh, J., Kim, M., 2020. JSI-GAN: GAN-based joint super-resolution and inverse tone-mapping with pixel-wise task-specific filters for UHD HDR video. Proc. AAAI Conf. Artif. Intell. 34 (7), 11287–11295.

Lanchantin, J., Wang, T., Ordonez, V., Qi, Y., 2020. General multi-label image classification with transformers. arXiv preprint arXiv:2011.14027.

Li, J., Fang, P., 2019. HDRNET: Single-image-based HDR reconstruction using channel attention CNN. In: Proceedings of the 2019 4th International Conference on Multimedia Systems and Signal Processing. In: ICMSSP 2019, pp. 119–124.

Li, H., Peers, P., 2017. CRF-net: Single image radiometric calibration using CNNs. In: Proceedings of the 14th European Conference on Visual Media Production (CVMP 2017). In: CVMP 2017, http://dx.doi.org/10.1145/3150165.3150170.

Liang, Z., Xu, J., Zhang, D., Cao, Z., Zhang, L., 2018. A hybrid l1-l0 layer decomposition model for tone mapping. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4758–4766. http://dx.doi.org/10.1109/CVPR.2018.00500.

Lin, S., Gu, J., Yamazaki, S., Shum, H.-Y., 2004. Radiometric calibration from a single image. In: Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition CVPR 2004, Vol. 2,. pp. II–938. http://dx.doi.org/10.1109/CVPR.2004.1315266.

Lin, S., Zhang, L., 2005. Determining the radiometric response function from a single grayscale image. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 2. pp. 66–73. http://dx.doi.org/10.1109/CVPR.2005.128, vol. 2.

Liu, K., Cao, G., Duan, J., Qiu, G., 2021. Lightness modulated deep inverse tone mapping. arXiv:2107.07907.

Liu, S., Huang, D., Wang, Y., 2018. Receptive field block net for accurate and fast object detection. In: European Conference on Computer Vision. pp. 404–419.

Liu, Y.L., Lai, W.S., Chen, Y.S., Kao, Y.L., Yang, M.H., Chuang, Y.Y., Huang, J.B., 2020. Single-image HDR reconstruction by learning to reverse the camera pipeline. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1648–1657. http://dx.doi.org/10.1109/CVPR42600.2020.00172.

Mangiat, J., Stephen, G., 2010. High dynamic range video with ghost removal. Proc. SPIE - Int. Soc. Opt. Eng. 7798, 30.

Mantiuk, R., Kim, K.J., Rempel, A.G., Heidrich, W., 2011. HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Trans. Graph. 30 (4), http://dx.doi.org/10.1145/2010324.1964935, URL http://resources.mpi-inf.mpg.de/hdr/vdp/.

Mantiuk, R.K., Myszkowski, K., Seidel, H.-P., 2015. High dynamic range imaging. In: Wiley Encyclopedia of Electrical and Electronics Engineering. American Cancer Society, pp. 1–42.

Marnerides, D., Bashford-Rogers, T., Hatchett, J., Debattista, K., 2018. ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. Comput. Graph. Forum 37 (2), 37–49.

Molchanov, P., Tyree, S., Karras, T., Aila, T., Kautz, J., 2017. Pruning convolutional neural networks for resource efficient inference. In: ICLR.

Nemoto, H., Korshunov, P., Hanhart, P., Ebrahimi, T., 2015. Visual attention in LDR and HDR images. In: International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM.

Pan, Z., Yi, X., Zhang, Y., Jeon, B., Kwong, S., 2020. Efficient in-loop filtering based on enhanced deep convolutional neural networks for HEVC. IEEE Trans. Image Process. PP (99), 1.

Raipurkar, P., Pal, R., Raman, S., 2021. HDR-cGAN: Single LDR to HDR image translation using conditional GAN. arXiv:2110.01660.

Shangzhe, W., iarui, X., Yu-Wing, T., Chi-Keung, T., 2018. Deep high dynamic range imaging with large foreground motions. In: Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part II. In: Lecture Notes in Computer Science, pp. 120–135.

Srikantha, A., Sidibé, D., 2012. Ghost detection and removal for high dynamic range images: Recent advances. Signal Process. Image Commun. 27 (6), 650–662.

Wang, X., Girshick, R., Mulam, H., He, K., 2018. Non-local neural networks. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7794–7803. http://dx.doi.org/10.1109/CVPR.2018.00813.

Wang, Z., Lu, Y., Li, W., Wang, S., Chen, X., 2021. Single image super-resolution with attention-based densely connected module. Neurocomputing 453, 876–884.

Xia, Y., Qin, T., Chen, W., Bian, J., Yu, N., Liu, T.-Y., 2017. Dual supervised learning. In: Proceedings of the 34th International Conference on Machine Learning - Volume 70. pp. 3789–3798.

Yan, Q., Gong, D., Shi, Q., Hengel, A., Zhang, Y., 2019. Attention-guided network for ghost-free high dynamic range imaging. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1751–1760. http://dx.doi.org/10.1109/CVPR.2019.00185.

Yang, X., Xu, K., Song, Y., Zhang, Q., Wei, X., Lau, R., 2018. Image correction via deep reciprocating HDR transformation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1798–1807.

Ye, Q., Xiao, J., man Lam, K., Okatani, T., 2021. Progressive and selective fusion network for high dynamic range imaging. arXiv:2108.08585.

Yi, Z., Zhang, H., Tan, P., Gong, M., 2017. DualGAN: Unsupervised dual learning for image-to-image translation. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 2868–2876.

Yu, H., Liu, W., Long, C., Dong, B., Zou, Q., Xiao, C., 2021. Luminance attentive networks for HDR image and panorama reconstruction. arXiv:2109.06688.

Zhang, J., Lalonde, J.F., 2017. Learning high dynamic range from outdoor panoramas. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 4529–4538.

Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 2242–2251.